Dick Reding's final copy 150

filename: detlevel/SM1030/2posavec.mdl

February 10, 1998

Personal comments submitted by Richard Reding, USEPA on "American Water Works Association - Review and Comment Letter Ballot - SMLB #107-20" attached to memorandum dated January 14, 1998 from Steven J. Posavec to Standard Methods Committee with a closing date of February 11, 1998.

Comments on 20th edition Section 1030, Data Quality

1030B, Measurement Uncertainty:

page 15, last line: An equation appears to be missing for the uncertainty statement.

1030C, Detection Level:

This section is a major rewrite of what had been 1030E (Method Detection Level) in the 19th edition. Section 1030E is renumbered as 1030C in the 20th edition. The 19th edition standard was a clear and concise descriptive standard. The proposed 1030C standard is a long and complicated debate about the relative merits of two ways (single concentration and calibration) to "design" a set of experiments to estimate a detection level. In addition to being difficult to comprehend, the discussion is not balanced, and as indicated in the attached comments, technically incorrect or incomplete in many ways. Both detection level designs have relative advantages and disadvantages; however, in proposed 1030C advantages of the calibration design are compared to disadvantages of the single concentration design. For example, no mention is made that the iterative single-concentration design requires less data for developing a detection or quantitation limit and considerably less data for verifying a detection or quantitation limit than the calibration design.

I recommend dropping the proposed standard and replacing it with the shorter and clearer 1030E standard from the 19th edition. If the Committee elects to propose a standard that is akin to a statistical debate, proposed 1030C should be extensively revised to correct technical errors and to provide a more balanced comparison and critique of the two detection level designs.

In the attachment is a more extensive critique of standard 1030 with an emphasis on proposed section 1030C.

ATTACHMENT: Comments on 20th edition Section 1030, Data Quality

The 19th edition 1030E standard impartially and briefly described four types of detection or quantitation estimates: IDL, LLD, MDL and LOQ. In 1030E the reader was given a brief overview of how the different estimates could be calculated from laboratory data. The standard did not specify a preferred estimate and referred readers to the literature for a thorough discussion of how to calculate and use these estimates. The 20th edition proposed section 1030C is an extensive and complicated discussion and comparison of two types of estimates: the single concentration design and the calibration design. No mention is made of the IDL or LLD. It is unclear from reading 1030C how an analyst would choose a procedure to estimate a detection level.

Other comments on proposed 1030E are referenced below by page number.

page 17, 4th from last line: Delete the word "only", because it demonstrates a bias.

page 17, last line and page 18, first 3 lines: The automobile speed limit analogy is not applicable. An analogy would be the ability of a speedometer to measure in the range of a few

5p

miles per hour or fractions of a mile per hour with some degree of confidence. The point at which to enforce a speed limit is policy established by a law-enforcement agency.

page 19, paragraph 2: If there is to be a discussion of Type II error, it would be appropriate to discuss Type I error on page 17. If discussed, they both need to be defined.

page 20, first 2 lines: The common probabilities are α=0.01; β=0.05 (see the ASTM 99/95 Interlaboratory Detection Estimate (IDE). If there is to be no discussion of these probabilities, the common ones should be used.

page 20, next to last paragraph. The statement about the controversial nature of the MDL should be moved to the Introduction section, 1030A. Detection or quantitation limits became controversial when use of the model of standard deviation vs. concentration was advanced. It would be prudent and helpful to analysts, managers and regulators to know that the MDL is widely used and has been codified in the *Code of Federal Regulations* (CFR) by USEPA in appendix B of 40 CFR part 136. Proposed section 1030C only cites a 1981 journal article as the authority for the MDL.

page 20, last paragraph: There is no mention of the iterative step in the MDL procedure. This iterative step can focus MDL measurements on the lowest concentration or amount that can be measured, which should be the objective of establishing a detection limit.

page 21, section b: The term "calibration design" creates confusion with the calibration process. An alternative terms are "Statistical model design," "Standard deviation model," or some other name that accurately represents the concept without confusion with existing analytical chemistry terms.

page 21, next to last paragraph, sentence beginning on line 2: Single-concentration designs assume that uncertainty is constant in the region of detection, an assumption that has been shown to be true for nearly all analytical systems.

page 21, next to last paragraph, sentence beginning on line 3: All designs that use data have this limitation, including the statistical model design, as evidenced by the example and figures 1030:3 - 1030:7, and as evidence by the data presented in Table 3 of Reference 18. The detection limit from all designs is dependent on the data used.

page 21, last 2 lines: The converse is also true; i.e., that the statistical model is more likely to overestimate variability and therefore detection levels because it is not focused on the lowest level that can be measured.

page 22, line 1: The primary disadvantage of the statistical model design is the cost, the additional effort required to gather sufficient data to allow a reliable estimate of standard deviation as a function of concentration, and the suggestion to purchase proprietary software to calculate the level.

page 22, paragraph 2: A reason for use of a prediction interval and its relevance to a detection limit are not given. Use of a prediction interval is arguable based on the ultimate use for the

detection limit.

page 22, last paragraph and page 23, paragraph 1: The use of different statistical models is a deterministic approach; i.e., the best model is fitted to the data. In a more recent paper (Anal. Chem. (1997) **69**, 3069-3075), a higher order model is used than those listed. However, if the correct model for standard deviation as a function of concentration is constant standard deviation at low levels transitioning to proportional standard deviation in the analytical range, a deterministic approach is incorrect because it fits data that may be in error. The data presented in Figures 1030:4 - 1030:7 demonstrate this problem. These are analytical system calibration data and were collected in the region of proportional standard deviation. Data should have been collected in a much lower region to address the detection capabilities of the analytical system. If data had been collected in the lower region, it is likely that the constant standard deviation model would have produced the best fit to the data.

page 22, section 3:
Data presented in the graphs of true and measured concentrations appear to be the same data presented in Figure 5.1 of Reference 12, with data added at 1.0 μg/L. These same data were presented in another paper (*Am. Env. Lab.* **8**, (1996) pp 6, 8, 10) with data added at 1, 0.5, 0.1, and 0.01 μg/L. The added data do not appear in the original data (Figure 5.1 of Reference 12), meaning that they were either added at a later date or were synthesized for illustrative purposes. Any examples of data to appear in Standard Methods should be real-world data, not synthesized examples, and should be part of a data set obtained as a result of a study directed at the issue being addressed, not as a part of testing directed at another purpose, in this case calibration. Use of data from two different data sets without mention is disingenuous. If data are being synthesized without mention of the synthesis, presentation of these data is scientific fraud.

page 23, section 3, and Table 1030:I:
Data used to establish the best model are not presented, so the reader does not know how to select the best model. Presumably, the model selected is based on residuals, but the residuals are not presented. Therefore, the reader is left with the impression that the way in which a detection limit is determined is to try various models and select the one that is lowest or closest to the reader's liking.

page 23, section 3, paragraph 1:
If a tolerance interval rather than a prediction interval is desirable, an explanation should be provided and the prediction interval should be dropped.

page 24, section 4:
The requirements in this section are onerous. Many organizations might not have (or wish to expend) the money and time to conduct a study with the extensive controls described in this proposed section. Simplifications should be suggested that will allow determination of a detection limit in a reasonable way. It is appears that the data presented in 1030:3 - 1030:7 were not gathered using the extensive requirements described in the proposed standard.

page 24, last paragraph, and page 25, paragraph 1, page 26:
Use of calibration data for a detection limit study contradicts principles of analytical chemistry, because calibration data are not used to estimate a detection level. Calibration data are used to

characterize an analytical system or instrument in the intended analytical range. To establish a detection limit, data must be gathered at levels below the calibration range; otherwise, the detection limit will be overstated. Further, although samples may be submitted blind, laboratories should be informed that it is expected that the analytes will be present in low or non-detectable concentrations and that the laboratories are to attempt to detected the analytes at as low a level as possible, in the same way that laboratories should be requested to address a low level environmental or other sample.

page 27, section b:
The definition for the LOQ is incorrect. LOQ = 10 s_0 (Anal. Chem. (1983) **55**, p 2217, column 1, paragraph 3).

page 28, paragraph 1:
It is stated that it may not be possible to obtain a valid estimate of s_0. However, nearly all detection and quantitation limits are based on hypothesis testing using s_0, including both the MDL and the "calibration design" statistical model design. When s_0 cannot be measured, it is common practice to use the standard deviation at the lowest concentration that can be measured.

page 28, paragraph 1, last line:
The LOQ cannot be negative if the proper definition of the LOQ is used (Anal. Chem. (1983) **55**, p 2217, column 1, paragraph 3).

page 28, section c, sentence 3:
Delete this sentence, because it shows an inappropriate and obvious bias against the MDL.

page 28, section c, sentences 4 and 5:
A bias against the second definition of the PQL appears.

page 29, word 1:
Change "the" to "some."

page 29, section e:
An advantage of the ML and LOQ are that they produce a quantitation limit regardless of the variability and heterscedaticity of the data. The ML/LOQ appear highly dependent on spike concentration because the author incorrectly uses calibration data for the purpose of establishing detection and quantitation limits. The proposed section 1030 ignores another feature of the definition of the ML, i.e. the ML is the lowest level at which the entire analytical system must provide an acceptable calibration point. In the presentation of the benzene data, this lowest calibration point would be 1 μg/L. However, in previous presentations (Figure 5.1 of Reference 12) these same data were limited to 4 μg/L, whereas in another paper (*Am. Env. Lab.* **8**, (1996) pp 6, 8, 10) these data were extended to 0.01 μg/L. If calibration data can be extended in this way and particularly if by synthesis, the arguments about the MDL/ML being dependent on spike concentration should be discounted, because they are not supported.

pages 29 and 30, section f:
It is stated that a problem with the ML/LOQ is that it may not always achieve 10% RSD whereas the AML does not guarantee 10% RSD at the AML. If neither achieve nor guarantee 10% RSD,

there is no advantage of one over the other on this point.

page 30, paragraph 1, sentence 3:
Inclusion of higher concentrations will further inflate the AML, resulting in an increased overestimate of the quantitation limit.

page 30, dagger:
Proprietary software? What connection does Scientific Software International have with the Standard Methods Committee? Are there public domain ways to effect these estimates?

1030D, Data Quality Objectives (DQOs):

Although the discussions of DQOs are interesting, the discussion is inadequate and somewhat afield from scope of Methods for the Examination of Water and Wastewater. The discussion might be more appropriate for environmental managers and decision makers, rather than laboratory personnel. However, if the committee believes that the discussion is useful, this comment should not be considered a negative.

###